

ASSIGNMENT #2: MISLEADING VISUALIZATION

Ivan Aristy

CS-GY 6313 B

1 Visualizations

The following are the two Visualizations that I have created for this Assignment, please note that the earnest visualization has elements that define its "earnestness", omitted as per the assignment's obfuscation mandate.

2 Report

2.1 Topic Selection

2.1.1 Topic Ideation

For the Misleading visualization I had a few ideas:

- Immigration and Crime Correlations: Showing an extrapolated correlation between immigration and crime rates could be made misleading by selecting data points that support a biased narrative. For example, not filtering by region, and actively sampling from regions with high crime rates.
- Ethnically Biased Drug Usage: Using the "US Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin" [2] dataset, we could use misleading bar charts to exaggerate the correlation between hispanic heritage and drug consumption. Additionally, by filtering out specific drugs we could make it especially misleading.
- Presidential Employment Claims: As much as I do not like Trump, which you'll see in the extra credit, the high employment rates during his presidency are largely do to the Covid-19 pandemic. We could use a line chart to show the employment rates over time, in contrast/comparison to Biden, and by not clarifying the context of the data, we could make it seem like Trump was solely responsible for the high employment rates.
- Immigration Crisis and Democratic Party Correlation: We could use a line graph to show the number of border crossings over time, and, by comparing Biden's and Trump's presidency, we could make it seem like the Democratic party is solely responsible for the immigration crisis. This could be further exaggerated by not showing the Covid-19 pandemic context, and/or picking legal border crossings to be also part of the data using the "Border Crossing Entry Data" [1] dataset.
- The Cost of Living vs Inflation: The relation between the American people's expenditures, home prices, income, and inflation rates are great for generating accurate and misleading content alike. For example, one could greatly overvalue the spending power of the American people by not accounting for inflation rates. Also, by sampling subsets of the population, we could make homes seem less or more affordable than they are.

2.1.2 Topic Selection and Question

I decided to go with the "The Cost of Living vs Inflation" topic.

The reason why is because cost of living and income statistics lend themselves to both areas of the spectrum of honesty. We can sample and transform numbers in multiple ways, incorrectly scale items, and/or dramatize the effects of specific economic factors to create misleading narratives on the issues americans care most about. This topic also facilitates my secondary goal for this assignment: Showing how changes in data selection and visualization techniques affect the perception of the viewer.

The question to answer is: *Has the increase in home prices made it harder for the average American to own a home?*

I believe that this question will be perfect for evaluating both ends of the spectrum.

We will first try to generate the correct visualization of the data, answering the question in the most clear and attentive way possible. Then, I aim to generate a visualization that creates confusion through various mediums. For example, the misleading visualization downplays, exaggerates, or eliminates important considerations like inflation, and/or misrepresents the increase in home prices through incorrect sampling.

2.2 Data

2.2.1 Required Data

We will focus on the analysis of yearly data.

The goal was to acquire data regarding:

- Home Prices: Real Home Prices, Non-Inflation Adjusted Home Prices, Home Price Index data, Rental Prices, Mortgage Prices, or Mortgage Rates.
- Income: Real Income, Non-Inflation Adjusted Income, Income Index data, or household earnings.
- Miscellaneous: American spending habits and/or miscellaneous expenditures.

2.2.2 Data Gathering

All data was obtained from the Federal Reserve Economic Data (FRED) database.

Here is a summary of the specific datasets acquired:

- **Mean Family Income**
 - **Series ID:** MAFAINUSA646N
 - **Units:** Current dollars
- **Nominal Median Household Income**
 - **Series ID:** MEHOINUSA646N
 - **Units:** Current dollars
- **Real Median Household Income**
 - **Series ID:** MEHOINUSA672N
 - **Units:** Constant dollars (adjusted for inflation)
- **Average Sales Price of Houses Sold**
 - **Series ID:** ASPUS
 - **Units:** Current Dollars
- **Median Sales Price of Houses Sold**
 - **Series ID:** MSPUS
 - **Units:** Current Dollars
- **Real Residential Property Prices**
 - **Series ID:** QUSR628BIS
 - **Units:** Index 2010=100
- **Housing Affordability Index**
 - **Series ID:** FIXHAI
 - **Units:** Index
- **Real Personal Consumption Expenditures**
 - **Series ID:** PCEC96
 - **Units:** Billions of Chained 2017 Dollars
- **Personal Consumption Expenditures**
 - **Series ID:** PCE
 - **Units:** Billions of Current Dollars
- **Retail Sales**
 - **Series ID:** RSAFS

- **Units:** Current dollars

2.3 Earnest Visualization: Iterative Improvement

2.3.1 Understanding The Data

For the earnest visualization, I want to accurately show whether or not it has become more difficult for the average american to own a home.

We have all the necessary data to answer this question, specifically, for plotting the real wages of the average american against the real home prices.

To eliminate all possible bias, we will use the Real Median Household Income and the Real Residential Property Prices datasets. By using the median of household income, we can avoid the bias of outliers, more accurately showing the average american, and, by using the real prices, we can avoid the bias of inflation.

2.3.2 Iterative Improvement

2.3.3 Visual Encodings

2.4 Misleading Visualization: Iterative Deceit

2.4.1 Data Obfuscation

2.4.2 Iterative Deceit

The first iteration had a focus on correctly plotting home prices vs real income:

2.4.3 Effectiveness

2.5 Final Designs

2.5.1 Do both answer the question?

2.5.2 Conclusion

References

- [1] *Border Crossing Entry Data*. Accessed: 2023-10-01. 2024. URL: <https://catalog.data.gov/dataset/border-crossing-entry-data-683ae>.
- [2] *Drug Overdose Death Rates by Drug Type, Sex, Age, Race, and Hispanic Origin: United States*. Accessed: 2023-10-01. 2024. URL: <https://catalog.data.gov/dataset/drug-overdose-death-rates-by-drug-type-sex-age-race-and-hispanic-origin-united-states-3f72f>.